Aalto University
School of Science

CS-E4660 Advanced Topics in Software Systems

Hands-on tutorial: Machine Learning pipeline on Edge

Minh Tri Nguyen
Ph.D student of Aalto University
Researcher at AaltoSEA

Who I am? and what is this tutorial about?

- **Who I am?**
  - I am Minh Tri Nguyen
  - MSc degree in Computer Science in 2019
  - PhD student at Aalto University
- **What is this tutorial about**
  - Machine Learning inference on edge devices.
  - A quick demo of deploying a simple ML pipeline on Raspberry Pi.

Overview

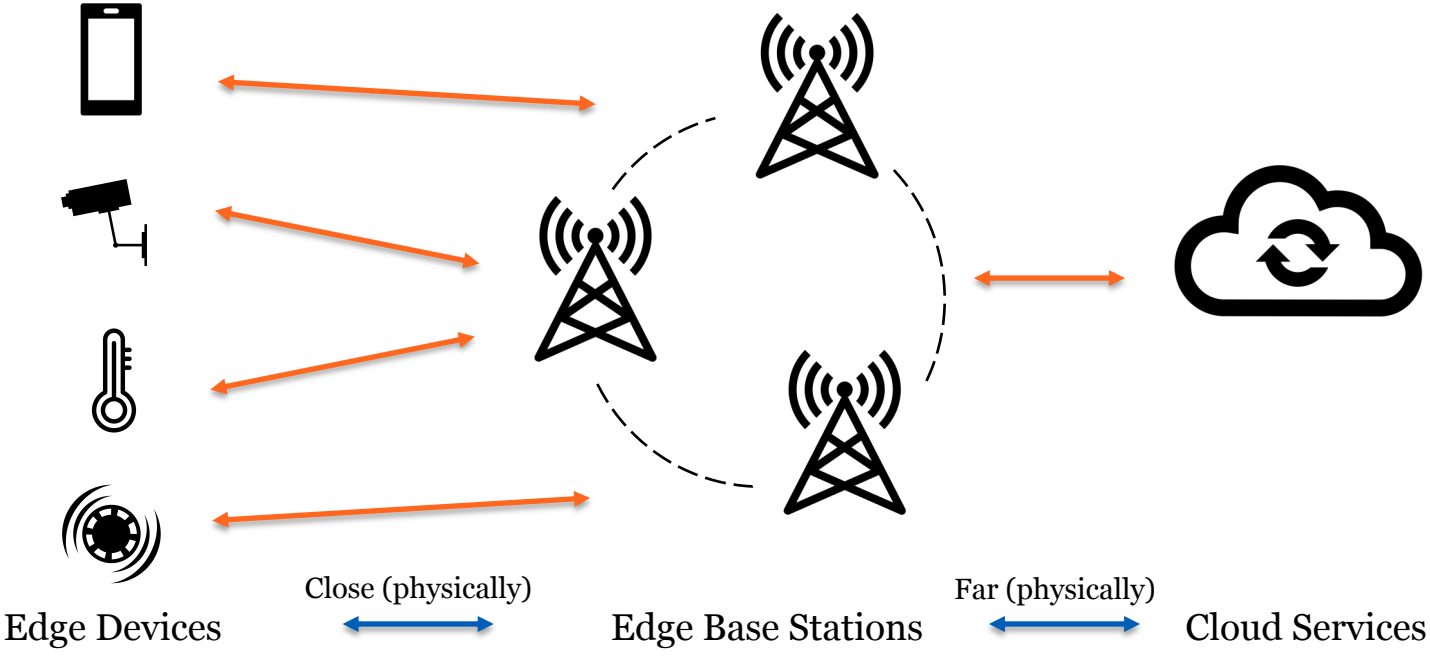
- **Why do we need to move ML to the edge**
  - Mitigating computing and network function burdens on the cloud as cloud resources are costly.
  - Edge devices are located close to user application/IoT devices/data sources, which is necessary for supporting real-time services.
  - …
- **Difficulties:**
  - Limited computing resources – CPU, memory, storage, accelerator, energy/power..
  - Network problems - unstable/reliable connectivity, low-bandwidth.
  - Low-level of supporting application development
  - Heterogeneous hardware
  - …

System Outline

- **IoT streaming data pipeline**
  - Sensor/Edge devices collecting data (e.g., environment monitoring data, video, image, …)
  - IoT Data Hub (MQTT broker, …)
  - Data ingestion service
- **Machine Learning pipeline**
  - Data ingestion, pre-processing, …
  - Model design, training, optimization
  - ML serving/inference (on the Edge)

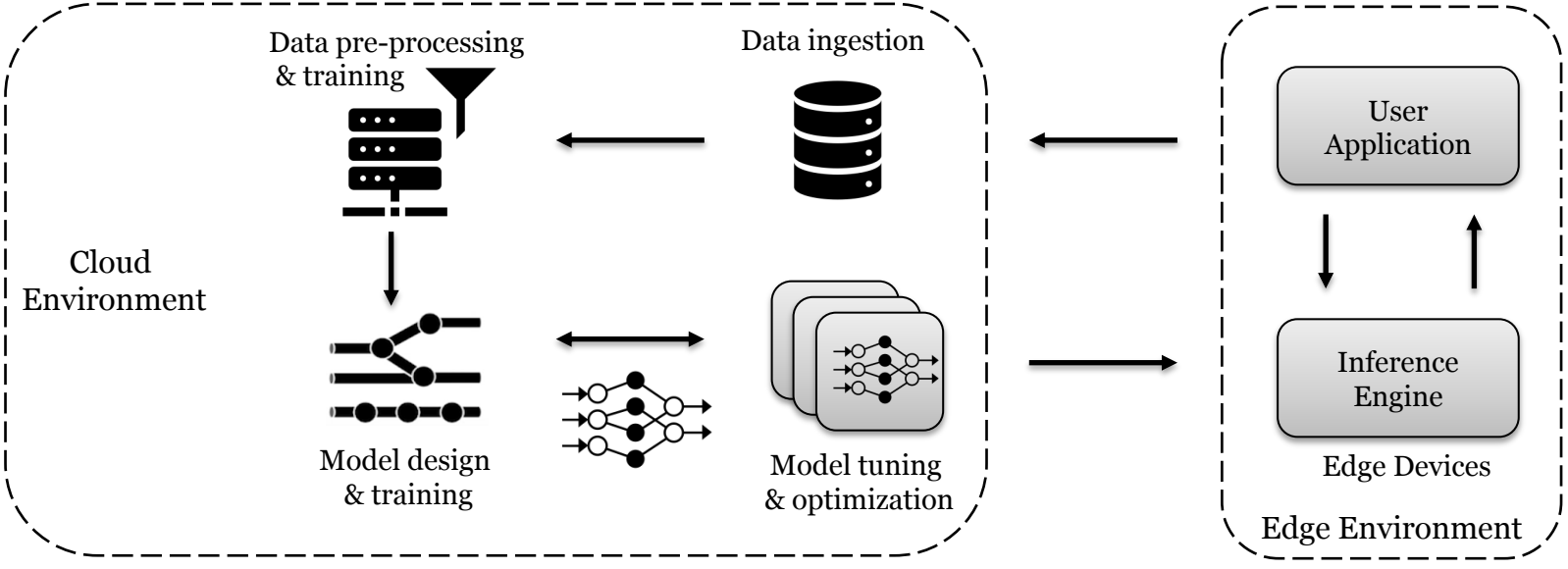
System Outline

- **IoT streaming data pipeline**
  - Implemented using MQTT: publisher, subscriber, ...



System Outline

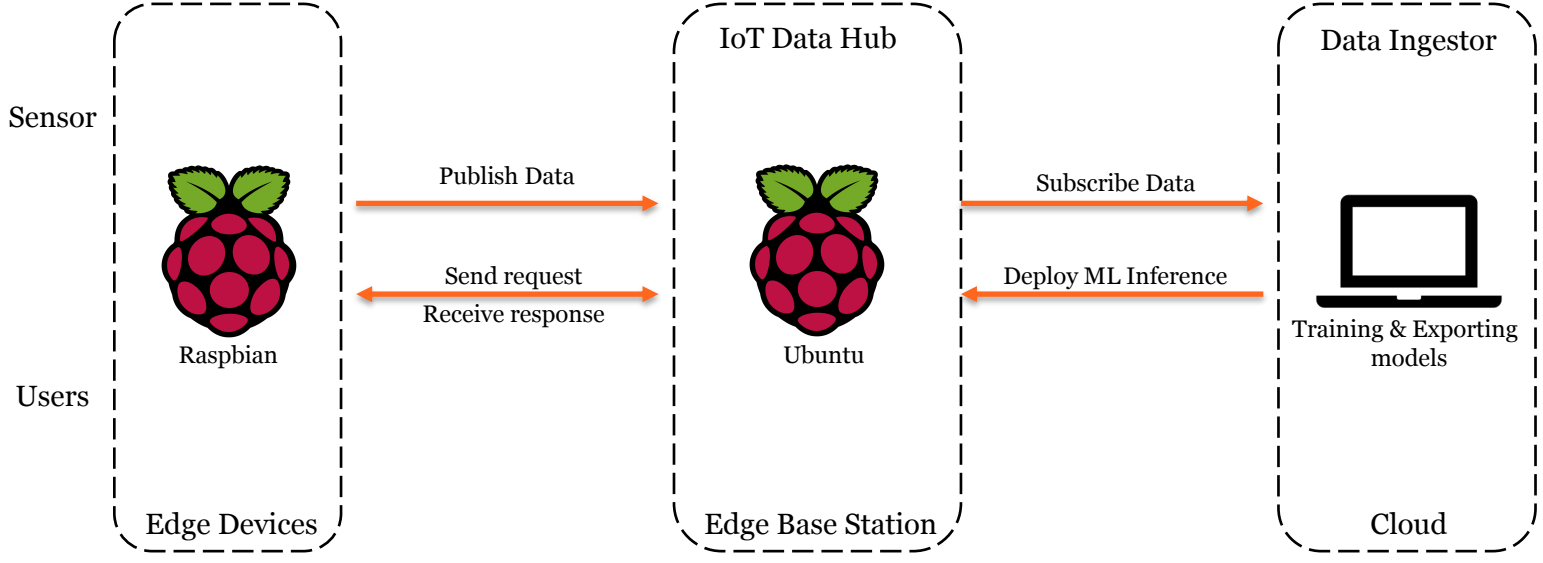
- **Machine Learning pipeline**



A Quick Guide for ML pipeline on Edge

- **Prerequisite**
  - Python 3 virtual environment
  - TensorFlow 2.3.0
  - TensorFlow Lite
  - Numpy
  - Java environment (JRE 11)
  - Pandas
  - Paho MQTT

A Quick Guide for ML pipeline on Edge



Discussion

Contact and Further information

- https://version.aalto.fi/gitlab/sys4bigml/cs-e4660
- https://docs.openvinotoolkit.org/2020.4/openvino_docs_MO_DG_Deep_Learning_Model_Optimizer_DevGuide.html

Email: tri.m.nguyen@aalto.fi